Segmentation of organs in pig offal using auto-context

Amaral, Telmo ; Kyriazakis, Ilias ; McKenna, Stephen; Plotz, Thomas

Published in:
2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)

DOI:
10.1109/ISBI.2016.7493511

Publication date:
2016

Document Version
Peer reviewed version

SEGMENTATION OF ORGANS IN PIG OFFAL USING AUTO-CONTEXT

Telmo Amaral[⋆] *Ilias Kyriazakis*[†] *Stephen J. McKenna*[‡] *Thomas Plötz*[⋆]

[⋆] Open Lab, School of Computing Science, Newcastle University, Newcastle upon Tyne, UK
[†] School of Agriculture, Food and Rural Development, Newcastle University, Newcastle upon Tyne, UK
[‡] CVIP, Computing, School of Science and Engineering, University of Dundee, Dundee, UK

ABSTRACT

The segmentation of 2D images of 3D non-rigid objects into their constituent parts can pose challenging problems, such as missing and occluded parts, weak constraints over the spatial arrangement of parts, and variance in form and appearance. These problems have been addressed via segmentation methods that incorporate spatial context information, such as the auto-context technique. In this paper, we address for the first time the problem of segmenting multiple organs in images of pig offal, a challenging image analysis task that constitutes an essential step towards automated screening at abattoir for signs of sub-clinical diseases. We applied auto-context segmentation to a large data set of images and explored the effect of complementing conventional context features with integral features suited to our application.

***Index Terms*—** Animal models and imaging, Internal organs, Image segmentation

1. INTRODUCTION

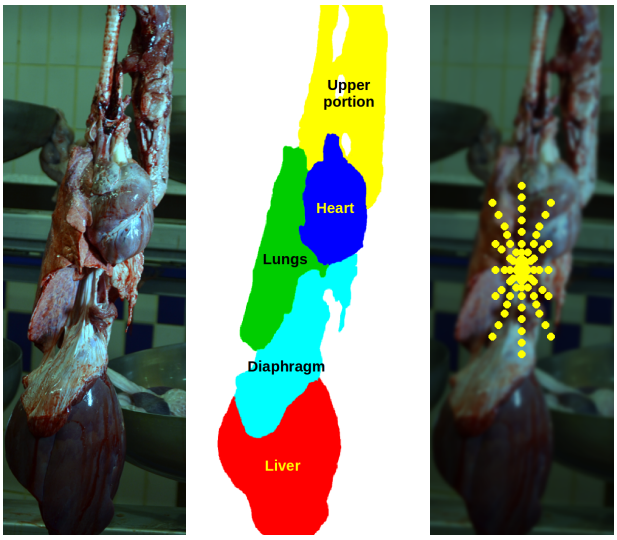
The segmentation of 2D images of 3D non-rigid objects into their multiple constituent parts can pose some challenging problems. In many biological and medical imaging settings, the spatial configuration of parts provides weak geometric constraints; parts vary in form and appearance, sometimes pathologically; parts can be missing; and viewpoint can be poorly controlled, leading to varied inter-part occlusions. Such problems have been addressed via the incorporation of spatial context data into segmentation techniques, by combining models such as conditional random fields (CRFs) [4] with inference algorithms like belief propagation (BP) [14]. Disadvantages common to many such techniques that aim to capture context information include their reliance on fixed spatial configurations with confined neighbourhood relations and the complexity of training procedures. The auto-context (AC) technique proposed by Tu and Bai [12] aims to address these limitations, by simultaneously integrating local image appearance and far-reaching context in an iterative algorithm that is easy to implement. An attractive feature of AC is that a prior "atlas" (obtained for example by averaging rigidly registered label maps created by experts) can be used as a source of contextual data for the initial iteration.

Visual inspection of carcasses is an important means of ensuring the safety and quality of meat products, enabling the detection of pathological conditions and public health hazards, as well as the provision of useful feedback to pig producers. However, manual inspection puts a strain on meat inspector resources, effectively limiting detailed screening for the purposes of health schemes, and exposes the inspection process to subjective assessment. This motivates the need for the development of automated meat inspection systems and is consistent with the wishes of the European Standards Agency to minimize handling of carcasses at abattoir.

Existing literature on multi-organ segmentation techniques focuses extensively on computer tomography (CT) images of human abdominal organs, reporting on a wide variety of techniques ranging from hierarchical atlases combined with statistical shape models [9] to level set optimisation [3]. Applications to meat production deal mostly with estimation of proportions of muscle, fat and bone from CT images, both *in vivo* and in carcasses, sometimes involving the segmentation of internal organs without distinguishing them individually [8, 2]. Work dealing with localisation of individual organs from photography or video is scarce and includes the segmentation of poultry spleen from surrounding viscera in the work of Tao et al. [11], as a stage in the automated detection of splenomegaly from ultraviolet and colour images. More recently, automatic recognition of ovine organs has been proposed by Stommel et al. [10] as a component of an envisaged robotic sorting system for sheep offal.

In this paper we introduce the demanding task of segmenting individual organs (namely the heart, lungs, diaphragm and liver) in colour images of non-digestive tract offal of pigs. This task constitutes an essential stage in the development of a wider system aimed at on-site screening for signs of subclinical diseases, which are characteristically organ-specific. Images of pig offal acquired at abattoir pose challenging problems, such as partially or totally missing organs, occlusions,

This project was co-funded by Innovate UK, the Biotechnology and Biological Sciences Research Council (BBSRC) and Tulip Ltd. It was carried out in collaboration with Tulip Ltd and Hellenic Systems Ltd. We thank Sheralyn Smith, Gareth Woodward and Lazar Valkov for the patient manual annotation of hundreds of images.



(a) (b) (c)

Fig. 1. (a) Example image of pig "pluck" and (b) corresponding manual annotation. (c) 91-point context "stencil" overlaid on example image.

severe deformation and lack of control over viewpoint. We employed AC based segmentation and report results on a large data set. In our experiments, we also investigated the effect of complementing conventional context features with two types of integral features designed for our application.

2. DATA AND PROBLEM

The data used in this work consisted of colour images of non-digestive tract offal obtained at an abattoir for pigs. We used 350 colour images of pig offal acquired with a single-lens reflex digital camera mounted on a tripod and LED lighting. Each image had a resolution of 3646×1256 pixels and showed a group of connected organs known as the "pluck", mainly comprising the heart, lungs, diaphragm and liver. These four organs were manually annotated on each image, with a fifth class label being used to mark the irrelevant portion of the pluck located above the heart and lungs (usually consisting of the trachea and tongue). Fig. 1(a) and 1(b) show an example of pluck and the corresponding manual annotation, where the upper portion, heart, lungs, diaphragm and liver are annotated in yellow, blue, green, cyan and red, respectively. More examples are shown in Fig. 2.

In this work it was assumed that the pluck had already been successfully segmented from the background (a relatively trivial task that can be accomplished based on hue and focus information). Thus, the problem at hand was that of classifying only foreground pixels into each of the five organ classes.

Algorithm 1 Training of auto-context model.

Given training set $S = \{(Y_j, X_j),\ j = 1..m\}$, obtain prior atlas $Q^{(0)}$ from label maps Y_j and use it to initialise probability maps $P_j^{(0)}$. For iteration $t = 1..T$:

1. Build a training set for the iteration, $S^{(t)} = \{(y_{ji},\ (X_j(N_i),\ P_j^{(t-1)}(i)),\ j = 1..m,\ i = 1..n)\}$
2. Train a classifier on image features extracted from $X_j(N_i)$ and context features extracted from $P_j^{(t-1)}(i)$.
3. Use the classifier to obtain new probability maps $P_j^{(t)}(i)$.

3. METHODS

3.1. Auto-context

The auto-context (AC) method, proposed by Tu and Bai [12], is an iterative pixel labelling technique. Label probabilities output at a given iteration are used as a source of contextual data that are concatenated with local image features, to form the input vector for the classifier used in the following iteration. For convenience, in the formal description that follows, we adopted a notation similar to that used by Tu and Bai.

Let S be a set of m training images X_j together with their label maps Y_j, i.e. $S = \{(Y_j, X_j),\ j = 1..m\}$. At each iteration t we want to train a classifier that outputs the probability distribution $p_{ji}^{(t)}$ over labels $y_{ji} \in \{1..K\}$ for pixel i in image X_j, given image patch $X_j(N_i)$ and label probability map $P_j^{(t-1)}(i)$. In $X_j(N_i)$, N_i denotes all pixels in the patch, and $P_j^{(t-1)}(i)$ is map $P_j^{(t-1)}$ output for image X_j at previous iteration $t-1$, but now centred on pixel i. In summary, we want a classifier that outputs the posterior probability $p_{ji}^{(t)}(y_{ji}|X_j(N_i),\ P_j^{(t-1)}(i))$.

The AC training procedure yields a sequence of classifiers, one per iteration, and is formally described in Algorithm 1. Before the first iteration, all probability maps $P_j^{(0)}$ can be initialised using a prior atlas $Q^{(0)}$, obtained by averaging all the training label maps Y_j, i.e. $Q^{(0)} = (\sum_j Y_j)/m$. On each iteration, given pixel i in image X_j, the feature vector input to the classifier is composed of local image features extracted from patch $X_j(N_i)$ concatenated with context features extracted from re-centered label probability map $P_j^{(t-1)}(i)$. Context features are typically the probabilities extracted from selected locations on map $P_j^{(t-1)}(i)$, including the central location that corresponds to pixel i itself.

Each of our AC models consisted of a series of multilayer perceptrons (MLPs), trained using Algorithm 1. Each MLP had one layer of hidden units with logistic activation and a softmax output layer. MLP training was based on a

regularised error $e_r = e + A \sum w^2$, obtained from cross-entropy error e and a regularisation term $A \sum w^2$ to penalise large weights w [1]. In our experiments, the number of hidden units was set to 20 (a value chosen with the help of 3-fold cross-validation) and A was set to 0.1. The NETLAB library [7] was used.

3.2. Local appearance features

We used local appearance features based on a multi-level Haar wavelet decomposition [6]. Each image was converted to the CIELUV colour space [5]. For each component (L*, u* and v*), the approximation wavelet coefficients, as well as the horizontal, vertical and diagonal squared detail coefficients, were obtained at three levels of decomposition. This resulted in 36 feature maps (3 image components $\times$ 4 wavelet coefficients $\times$ 3 levels of decomposition), all rescaled to match the original dimensions of the image.

We then sub-sampled feature maps and label maps by a factor of 20 along both dimensions, which yielded maps with 180$\times$60 points. This was found to provide sufficient detail for our purposes. As explained in Section 2, we were concerned only with points within the foreground region of each image (i.e. within the pluck). On average, approximately 5,700 points per image belonged to the foreground. For each point we had a vector of 36 feature values and a class label.

3.3. Context features

For each point on a sub-sampled image, context features were extracted for the point itself and for 90 surrounding points, as defined by a sparse star-shaped "stencil" illustrated in Fig. 1(c). On the first iteration of AC, context features for each image point consisted of the 5 class probabilities provided by the prior atlas on all 91 context points. On the second and subsequent iterations, context features consisted of the 5 class probabilities output by the classifier on the previous iteration, again on all 91 context points. This yielded 91$\times$5=455 context features per image point.

Optionally, we complemented the context features described above with two types of integral context features suitable for our application. The relative positions of organs along the vertical direction vary little from image to image, given that each pluck hangs from a hook and the part of the pluck that is attached to the hook is very consistent across plucks. Thus, given a point on an image, class probabilities averaged over the *row* to which the point belongs provide the classifier on the next iteration with useful information as to which organs are likely to occur at that particular height. For example, a row containing heart is likely to contain also lungs, but very unlikely to contain liver.

In contrast, the relative positions of organs along the horizontal direction vary considerably from image to image, given that we had no control over the orientation of the pluck around the vertical axis. The heart, in particular, is sometimes fully occluded. Nevertheless, organs are fairly consistent in size from pig to pig. Thus, class probabilities averaged over the *whole* image reflect the proportions of the pluck covered by each visible organ, and provide the next classifier with useful information on which organs are likely to be visible and how visible they are. For example, a small proportion of visible diaphragm is consistent with a hidden heart and a corresponding large proportion of lung. When incorporating the two types of integral features described above, we extracted for each point a total of 455+2$\times$5=465 context features.

4. EXPERIMENTAL VALIDATION

4.1. Cross-validation strategy

The 350 available images were divided into 10 subsets of 35 images each, and 10-fold cross-validation experiments were carried out to assess the performance of AC with and without integral context features.

On each fold, there were 35$\times$9=315 training images. Given that each image had on average 5,700 foreground points, we had approximately 315$\times$5,700$\simeq$1,800,000 training samples available per fold. From these, 1600 samples were randomly picked for each of the 5 classes, resulting in a balanced set of 1,600$\times$5=8000 training samples per fold. This random sub-sampling was done for computational reasons, to limit the time needed to train the MLPs. Each training sample consisted of a local feature vector, a context feature vector and a class label. The training samples collected on each fold were used to train an AC model via 5 iterations of Algorithm 1.

On each fold, there were 35 test images. Given that each image had on average 5,700 foreground points, we had approximately 35$\times$5,700$\simeq$200,000 test samples available per fold. Within each fold, the trained AC model was tested on all 200,000 test samples.

4.2. Results and discussion

Table 1 shows the means and standard deviations of several performance metrics computed at the end of the 5th AC iteration over 10 folds, both without and with integral context features. The metrics used were: accuracy, mean class accuracy (MCA), adjusted Rand index (ARI) [13], and mean class F-score (MCF). Paired fold-wise results (with and without integral context) were compared via two-tailed Student's t-tests, yielding p-values that ranged from 0.0052 to 0.0526.

Fig. 3 plots the mean value of the adjusted Rand index over all five iterations, with and without integral context. On iteration 1, both methods shared the same result, as context features were extracted from the same prior atlas. The evolution of each method's performance then followed the typical pattern reported by Tu and Bai [12]: the largest improvement occurred on the 2nd iteration and performance nearly levelled

Table 1. Mean (standard deviation) of several performance metrics computed after the 5th AC iteration, over 10 folds, without and with integral context features.

Integral context	Accuracy	MCA	ARI	MCF	Class F-score				
					Upper	Heart	Lungs	Diaph.	Liver
Without	0.890 (0.017)	0.956 (0.007)	0.772 (0.030)	0.869 (0.021)	0.921 (0.013)	0.757 (0.055)	0.871 (0.023)	0.840 (0.021)	0.956 (0.011)
With	0.895 (0.015)	0.958 (0.006)	0.781 (0.026)	0.875 (0.018)	0.925 (0.014)	0.778 (0.046)	0.878 (0.025)	0.838 (0.024)	0.957 (0.007)

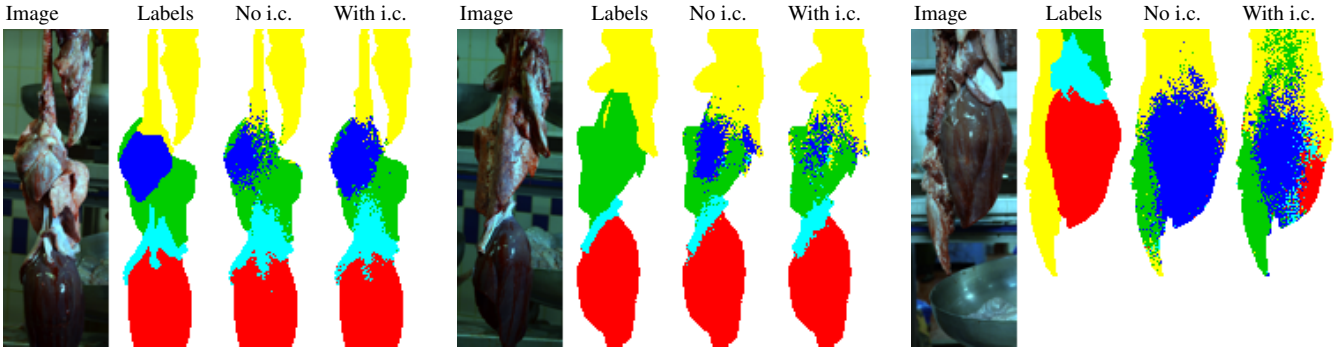


Fig. 2. Ground-truth labels, segmentation results without integral context (i.c.), and segmentation results with integral context, for three example images. (Best viewed in colour.)

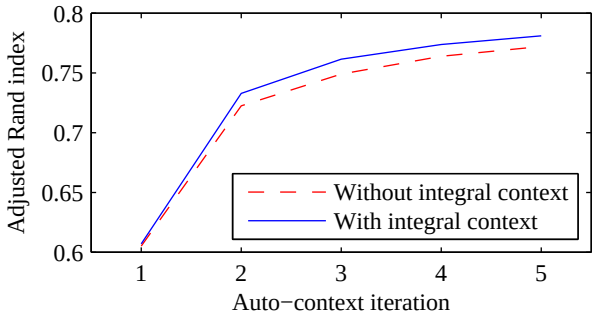


Fig. 3. Evolution of mean adjusted Rand index over five AC iterations, without and with integral context features.

off on the 5th iteration. It is interesting to note that integral context features had a beneficial effect on performance, despite corresponding to only 10 features out of 465.

Also shown in Table 1 are class-specific F-scores. The heart registered the lowest F-scores, which is not surprising. Being relatively small, this is the organ whose 2D projection on each image is most affected by the orientation of the pluck: it can be fully visible at the centre, fully or partially visible on either side of the pluck, or completely hidden. The heart also registered the largest improvement in F-score with the addition of integral context, suggesting that integral features help to deal with the unpredictability of the heart's presence and position in the image. In other words, integral context helped to deal with multi-modality in the data.

For three test images, Fig. 2 shows the ground truth labelling and the segmentation results obtained without and with integral context on the 5th iteration. Note that a simple denoising post-processing step would have improved the quality of segmentation results, but we left that step out to more clearly show the effect of adding integral context.

The importance of integral features is most visible in cases like that of Fig. 2(a), in which standard (stencil based) context was not enough to yield a confident segmentation of the heart. Fig. 2(b) illustrates the reverse situation, where integral features helped to dissipate a mistakenly segmented heart. In this case, the integral features representing class probabilities averaged over the whole image will have reflected the small area occupied by the diaphragm and large area covered by the liver, thus helping to identify a pluck whose dorsal aspect faced the camera, hiding the heart. Nevertheless, in rare situations where the pluck hangs in an unusual way, as in the example shown in Fig. 2(c), severe segmentation problems may occur that integral features are unable to correct. This suggests the need for more training examples of this kind.

5. CONCLUSION

We introduced the task of organ segmentation in the context of visual inspection of carcasses. We tackled this problem via auto-context segmentation and investigated the effect of complementing stencil-based context features with integral features suited to our application. Results on a large data set of pig offal images were reported, showing statistically significant improvements with the introduction of integral features. Future work will focus on dealing more explicitly with the multi-modal nature of the data, noticeable especially in the high variability of heart labels.

6. REFERENCES

[1] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[2] L. Bünger, C. Glasbey, G. Simm, J. Conington, J. Macfarlane, K. McLean, K. Moore, and N. Lambe. *CT Scanning - Techniques and Applications*, chapter Use of X-ray computed tomography (CT) in UK sheep production and breeding, pages 329–348. InTech, 2011.

[3] T. Kohlberger, M. Sofka, J. Zhang, N. Birkbeck, J. Wetzl, J. Kaftan, J. Declerck, and S. Zhou. Automatic multi-organ segmentation using learning-based segmentation and level set optimization. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 338–345, 2011.

[4] S. Kumar and M. Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1150–1157, 2003.

[5] M. Mahy, L. Eycken, and A. Oosterlinck. Evaluation of uniform color spaces developed after the adoption of CIELAB and CIELUV. *Color Research & Application*, 19(2):105–121, 1994.

[6] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[7] I. Nabney. *NETLAB: algorithms for pattern recognition*. Springer, 2002.

[8] E. Navajas, C. Glasbey, K. McLean, A. Fisher, A. Charteris, N. Lambe, L. Bünger, and G. Simm. In vivo measurements of muscle volume by automatic image analysis of spiral computed tomography scans. *Animal Science*, 82(04):545–553, 2006.

[9] T. Okada, K. Yokota, M. Hori, M. Nakamoto, H. Nakamura, and Y. Sato. Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 502–509, 2008.

[10] M. Stommel, W. Xu, P. Lim, and B. Kadmiry. Robotic sorting of ovine offal: Discussion of a soft peristaltic approach. *Soft Robotics*, 1(4):246–254, 2014.

[11] Y. Tao, J. Shao, K. Skeeles, Y. Chen, et al. Detection of splenomegaly in poultry carcasses by UV and color imaging. *Transactions of the ASAE-American Society of Agricultural Engineers*, 43(2):469–474, 2000.

[12] Z. Tu and X. Bai. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1744–1757, 2010.

[13] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):929–944, 2007.

[14] J. Yedidia, W. Freeman, and Y. Weiss. Generalized belief propagation. In *Neural Information Processing Systems (NIPS)*, volume 13, pages 689–695, 2000.